

# ROBUST BAYESIAN EXPERIENCE RATING

BY

RENÉ SCHNIEPER

## ABSTRACT

Different rating methods which allow for exceptional large claims are discussed. A robust Bayesian statistical model is proposed which can cope with non negative, skewed data. An example from fire insurance is analyzed. The performance of the posterior mean is compared to the performance of a robust credibility estimator.

## KEYWORDS

Experience rating, Bayesian statistics, robust Bayesian statistics, credibility, robust credibility, Markov Chain Monte Carlo

## 1. INTRODUCTION

The main advantage of Bayesian statistical methods is to allow the use of a priori information in a coherent way. This is important in an insurance context where in addition to the data there is usually a larger quantity of collateral or of a priori information available. It is essential in a reinsurance context where large claims play an important role and where by definition the amount of data is scarce. In such a context Bayesian techniques are especially powerful (see e.g. R. Schnieper (1993)).

Bayesian statistical models are sometimes criticized for their alleged lack of robustness. In order to keep the models tractable, practitioners often use distributions which have natural conjugate priors and are thus closed under sampling: the posterior density is of the same type as the prior density. The information conveyed by the data only leads to an update of the parameters of the prior density. (Examples are found for instance in J. Aitchison and I.R. Dunsmore (1975).) When it comes to computing the posterior mean, the most commonly used such likelihoods have indeed the property that they do not weigh down 'large' outlying observations thus producing estimates which are not robust. Credibility estimators can be viewed as another attempt to keep the estimation of the posterior mean mathematically tractable. They are least squares estimates of the posterior mean which are linear in the observations. As such they cannot weigh down 'large' outlying observations.

This lack of robustness however is not inherent to Bayesian statistics but is the consequence of the choice of the statistical model i.e. of the likelihood and of the prior distribution. We propose statistical models which can cope with large outlying observations. These models lead to posterior means which have to be computed numerically. This increased mathematical complexity is compensated by the fact that the models provide an automatic mean of detecting and accommodating aberrant observations or extraordinary large claims.

We start by giving a short presentation of the general Bayesian framework and of credibility theory. We look at different models which provide both an unbiased estimator of the pure risk premium and a robust treatment of large claims. It is seen that depending on the model, large claims are truncated or rejected. All our models are scale mixtures of normal distributions.

Claims data often have a skewed distribution and cannot be properly modelled by scale mixtures of normal. We show that many of our results also apply to properly reparametrized skewed distributions. The case of the reparametrized gamma distribution is analyzed.

The present paper should be viewed in the broader context of robust Bayesian analysis. For an overview of the topic see e.g. Berger (1993). We only consider the case of robustness in respect of the likelihood. Robustness in respect of prior information can be achieved by applying similar techniques to the prior distribution. We do not treat robustness in respect of the utility function since in an insurance context we are primarily interested in the pure risk premium, i.e. the posterior mean of the data, and hence we want to minimize the mean square error. We focus on fat-tailed distributions. We have not analyzed noninformative and partially informative priors as a mean to obtain robust statistical procedures. Non parametric Bayes procedures are discussed. Special cases of such procedures — data trimming and, more generally, the use of credibility techniques applied to robust estimators of the mean (see Gisler (1980), Künsch (1992) and Gisler, Reinhard (1993)) — are especially worth mentioning. In the example of section 4 we compare the performance of such an estimator with the performance of a fully fledged Bayesian analysis.

In general the statistical procedures we propose consist of a simultaneous sampling of the mean and of the variance.

## 2. GENERAL FRAMEWORK

### 2.1. Bayesian Model

Let  $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$  be a random sample from a common distribution with an unknown — possibly vector valued — parameter  $\theta$ . The unknown parameter  $\theta$  is treated as a random variable. Given  $\theta = \theta$ , the  $\tilde{x}_i$  are independent with common density  $f(x|\theta)$ , which is referred to as the likelihood of  $\tilde{x}_i$ . In an insurance context the  $\tilde{x}_i$ 's usually denote claims. As is customary in bayesian statistics, densities are indexed by their arguments. Thus  $f(x)$  and  $f(\theta)$  may be different densities. A generalization of the following results to the case where the probability distributions are not differentiable is straightforward and is

omitted here. The possible values of  $\tilde{\theta}$  are given by a density  $f(\theta)$ , the prior density. Bayes' theorem states that given a realization of  $\underline{x} = (x_1, x_2, \dots, x_n)$ , the posterior density of  $\tilde{\theta}$  is

$$f(\theta | \underline{x}) = \frac{f(\underline{x} | \theta) f(\theta)}{f(\underline{x})}$$

where

$$f(\underline{x} | \theta) = \prod_{i=1}^n f(x_i | \theta)$$

and

$$f(\underline{x}) = \int f(\underline{x} | \theta) f(\theta) d\theta$$

does not depend on  $\theta$ .

Thus

$$f(\theta | \underline{x}) = k f(\underline{x} | \theta) f(\theta)$$

where  $k$  is a constant independent of  $\theta$ .

Often one is interested in some function of  $\theta$ , e.g. in the pure risk premium

$$m(\theta) = E(\tilde{x} | \theta) = \int x f(x | \theta) dx.$$

A posteriori, i.e. after one has observed a realization of  $n$  claims  $\underline{x} = (x_1, \dots, x_n)$  the estimate of  $m(\theta)$  which minimizes the mean square error is

$$m(\underline{x}) = E(m(\tilde{\theta}) | \underline{x}) = \int m(\theta) f(\theta | \underline{x}) d\theta.$$

The predictive density

$$f(y | \underline{x}) = \int f(y | \theta) f(\theta | \underline{x}) d\theta$$

is a further quantity of interest.

### Example

Let  $\tilde{x}_1, \dots, \tilde{x}_n$  be normally distributed with unknown mean  $\theta$  and known precision  $p$  (the precision is the multiplicative inverse of the variance)

$$f(\underline{x} | \theta) = \prod_{i=1}^n f(x_i | \theta) = \left( \frac{p}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{p}{2} \sum_{i=1}^n (x_i - \theta)^2}.$$

The prior distribution of  $\tilde{\theta}$ , the unknown sample mean, is itself normal with known mean  $m$  and known precision  $q$

$$f(\theta) = \sqrt{\frac{q}{2\pi}} e^{-\frac{q}{2}(\theta - m)^2}$$

According to Bayes' theorem, the posterior density is

$$\begin{aligned} f(\theta | \underline{x}) &= k f(\underline{x} | \theta) p(\theta) \\ &= k e^{-\frac{1}{2} (p \sum_{i=1}^n (x_i - \theta)^2 + q(\theta - m)^2)} \\ &= k e^{-\frac{np+q}{2} (\theta - \frac{np \cdot \bar{x} + qm}{np+q})^2}. \end{aligned}$$

And it is seen that the posterior distribution of  $\tilde{\theta}$  is again a normal distribution, i.e. the normal likelihood with a normal prior is closed under sampling.

The posterior mean is

$$E(\tilde{\theta} | \underline{x}) = \frac{np \cdot \bar{x} + qm}{np + q}$$

The posterior precision is

$$\text{Var}^{-1}(\tilde{\theta} | \underline{x}) = n \cdot p + q$$

i.e. it is the sum of the precisions of the observations and of the prior density.

The mean square error of the posterior mean is

$$\begin{aligned} E(E(\tilde{\theta} | \underline{x}) - \tilde{\theta})^2 &= E[E(\tilde{\theta} - E(\tilde{\theta} | \underline{x}))^2 | \underline{x}] \\ &= E(\text{Var}(\tilde{\theta} | \underline{x})) \\ &= \frac{1}{np + q} \end{aligned}$$

Whereas the mean square error of the classical estimator is

$$E(\bar{X} - \tilde{\theta})^2 = E_{\theta}(\text{Var}(\bar{X} | \theta)) = \frac{1}{np}$$

Hence

$$E(E(\tilde{\theta} | \underline{x}) - \tilde{\theta})^2 = E(\bar{X} - \tilde{\theta})^2 \cdot \frac{n}{n + \frac{q}{p}}.$$

If  $n$  is small compared to the ratio of precisions  $q/p$ , the Bayesian estimator is much more precise than the classical estimator.

The predictive density is

$$f(y|\underline{x}) = k \cdot \int e^{-\frac{p}{2}(y-\theta)^2} e^{-\frac{np+q}{2}(\theta-m(\underline{x}))^2} d\theta$$

where

$$m(\underline{x}) = E(\tilde{\theta}|\underline{x}) = \frac{np\bar{x} + qm}{np + q}.$$

After some straightforward simplifications we obtain

$$f(y|\underline{x}) = k \cdot e^{-\frac{p}{2} \frac{np+q}{(n+1)p+q} (y-m(\underline{x}))^2}$$

and the predictive distribution is normal with expectation  $m(\underline{x}) = E(\tilde{\theta}|\underline{x})$  and with precision  $Var^{-1}(y|\underline{x}) = p \cdot \frac{np+q}{(n+1)p+q}$

## 2.2. Credibility Theory

Let  $\tilde{x}_1, \dots, \tilde{x}_n$  be a random sample characterized by an unknown risk parameter  $\theta$ . Given  $\theta$ ,  $\tilde{x}_1, \dots, \tilde{x}_n$  are independent with finite second moments

$$\begin{aligned} E(\tilde{x}_i|\theta) &= m(\theta) \\ Var(\tilde{x}_i|\theta) &= v(\theta). \end{aligned}$$

The pure risk premium  $m(\theta)$  is to be approximated by a premium which is linear in the observations

$$\widehat{m(\theta)} = \alpha_0 + \sum_{j=1}^n \alpha_j x_j$$

and which minimizes the mean square error

$$E\left(\widehat{m(\theta)} - m(\theta)\right)^2 = \min!_{\alpha_0, \alpha_1, \dots, \alpha_n}$$

It is easily seen that the optimal linear premium is a weighted average of the individual mean and of the a priori mean

$$\widehat{m(\theta)} = z\bar{x} + (1-z)m$$

where

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ m &= E(m(\tilde{\theta})) \end{aligned}$$

The weight  $z$  given to the individual mean  $\bar{x}$  is called the credibility factor. It is equal to

$$z = \frac{n \cdot \text{Var}(m(\tilde{\theta}))}{n \cdot \text{Var}(m(\tilde{\theta})) + E(\text{Var}(\bar{x}|\theta))}$$

In the notation of the preceding example we have

$$\text{Var}(m(\tilde{\theta})) = q^{-1} \quad E(\text{Var}(\bar{x}|\theta)) = p^{-1}$$

and it is seen that in the case of a normal likelihood with known variance and of a normal prior the credibility formula is equal to the a posteriori mean. This result generalizes to simple exponential families where the mean is the sufficient statistic and where the prior is the natural conjugate prior (see W.S. Jewell, (1974)).

The mean square error of the credibility formula is

$$E(\widehat{m(\theta)} - m(\theta))^2 = \text{Var}(E(\bar{x}|\theta)) \cdot (1 - z).$$

The mean square error of the classical estimator is

$$E(\bar{x} - m(\theta))^2 = \frac{p^{-1}}{n} = \frac{1}{n} E_{\theta}(\text{Var}(\bar{x}|\theta))$$

Hence

$$E(\widehat{m(\theta)} - m(\theta))^2 = z \cdot E(\bar{x} - m(\theta))^2$$

From the above it is seen that for  $n$  small compared to the ratio of within and between variances  $E(\text{Var}(\bar{x}|\theta))/\text{Var}(m(\tilde{\theta}))$ , the Bayesian estimator is much more precise than the classical estimator.

The general Bayesian framework and the credibility theory presented here can be generalized to different cohorts of risks, to observations with heterogeneous weights, to general linear models, etc.

### 3. ROBUSTNESS

An estimator is robust if it can cope with ‘large’ outlying observations. In an insurance rate making context it is usually desirable that exceptional large claims do not unduly influence the rate charged to a given risk. At the same time it is essential, that the proposed estimators are biasfree, since insurance rates must be such, that the premium is sufficient to also cater for the exceptional large claims.

### 3.1. Bayesian Models

The posterior mean in the example of section 2.1 is a non-robust estimator since

$$\lim_{x_n \rightarrow \infty} E(\tilde{\theta} | \underline{x}) = \infty$$

i.e. one ‘large’ claim fully determines the premium charged to a given risk. It is well known, that replacing the normal likelihood by a  $t$ -likelihood leads to a posterior mean which discards extreme observations.

$$\lim_{x \rightarrow \pm \infty} E(\tilde{\theta} | x) = E(\tilde{\theta}).$$

For a proof see O’Hagan (1994). In the case of a Laplace (double-exponential) likelihood the posterior mean is also a robust estimator with extreme observations being truncated

$$\lim_{x \rightarrow \pm \infty} E(\tilde{\theta} | x) = E(\tilde{\theta}) \pm \frac{\sqrt{2q}}{p}$$

where  $p$  and  $q$  are the within and between variance components as defined in section 2.2. For a proof of the statement see appendix 1.

Both the  $t$ -distribution and the Laplace distribution are scale mixtures of normal distributions. The Laplace distribution is obtained as a mixture of normal distributions with the variance being exponentially distributed. The  $t$ -distribution is obtained as a mixture of normal distributions with the precision — the multiplicative inverse of the variance — being  $\chi^2$ -distributed. Sampling from a Laplace distribution (or a  $t$ -distribution) can thus be viewed as sampling from a normal distribution with unknown variance. (unknown precision). Extreme observations are assumed to come from a distribution with a large variance (small precision) and get truncated (discarded).

The above mentioned models — Laplace respectively  $t$ -likelihood with a normal prior — are not appropriate to model non-life claims data which are non-negative and usually skewed. The models however hint at the way to robustify non-negative skewed distributions.

Let  $x$  denote some piece of claims information — a loss ratio, a burning cost, a claims intensity, etc. — pertaining to a given risk in a given period. Let  $\mu$  denote the unknown expected value of  $\tilde{x}$  and  $\tau$  the unknown variance of  $\tilde{x}$ . For a suitable choice of the densities  $f(x|\mu, \tau)$  and  $f(\tau|\mu)$  the likelihood

$$f(x|\mu) = \int f(x|\mu, \tau) f(\tau|\mu) d\tau$$

and the prior  $f(\mu)$  lead to a posterior mean

$$E(\tilde{\mu} | x) = \frac{\int \mu f(x|\mu) f(\mu) d\mu}{f(x)}$$

which is robust.

The Gamma density is a convenient choice of skewed density. It is usually parametrized in the following way

$$f(x|\gamma, c) = \frac{c^\gamma}{\Gamma(\gamma)} e^{-cx} x^{\gamma-1}$$

with  $E(\tilde{X}|\gamma, c) = \frac{\gamma}{c}$  and  $Var(\tilde{X}|\gamma, c) = \frac{\gamma}{c^2}$ .

Equating  $\mu = \frac{\gamma}{c}$  and  $\tau = \frac{\gamma}{c^2}$  we obtain the following reparametrization

$$f(x|\mu, \tau) = \frac{\left(\frac{\mu}{\tau}\right)^{\frac{\mu^2}{\tau}}}{\Gamma\left(\frac{\mu^2}{\tau}\right)} e^{-\frac{\mu}{\tau}x} x^{\frac{\mu^2}{\tau}-1}$$

We also assume that  $\mu$  and  $\tau$  are independent and are themselves gamma distributed

$$\begin{aligned} f(\mu) &= k \cdot e^{-\frac{m}{b}\mu} \cdot \mu^{\frac{m^2}{b}-1} \\ f(\tau) &= k \cdot e^{-\frac{w}{t}\tau} \cdot \tau^{\frac{w^2}{t}-1} \end{aligned}$$

Obviously the likelihood of  $\mu$

$$f(x|\mu) = \int f(x|\mu, \tau) f(\tau) d\tau d\mu$$

is no longer available in analytical form.

The posterior mean

$$E(\mu|x) = \frac{\int \mu f(x|\mu, \tau) f(\tau) f(\mu) d\tau d\mu}{f(x)}$$

can nevertheless be obtained through numerical integration.

It is no longer possible to derive the general behavior of  $\lim_{x \rightarrow \infty} E(\mu|x)$ . However in the case of the example given in appendix 2 it is seen that the posterior mean is robust.

In practical rate making applications, we usually have more than one piece of claims data. The design of experiment is usually such that we have  $N$  groups of claims related data. Group number  $i$  consists of  $n_i$  different observations

$$x_{i1}, x_{i2}, \dots, x_{in_i}$$

with a common unknown mean  $\mu_i$  and unknown variances  $\frac{\tau_{ij}}{v_{ij}}$ . The  $v_{ij}$  are typically known volume measures, such as premium incomes, in the case of loss



ratios or sums insured in the case of claims intensities. It is assumed that given the means and the variances, the observations are independent. The likelihood is thus

$$\prod_{i=1}^N \prod_{j=1}^{n_i} f(x_{ij} | \mu_i, \tau_{i,j})$$

It is also assumed that the parameters pertaining to different groups of claims data are independent thus

$$E(\mu_i | D) = \frac{\int \mu_i \cdot \prod_{j=1}^{n_i} f(x_{ij} | \mu_i, \tau_{ij}) f(\mu_i, \tau_{i1}, \dots, \tau_{in_i}) d\mu_i d\tau_{i1}, \dots, d\tau_{in_i}}{f(x_{i1}, \dots, x_{in_i})}$$

where  $D$  is the set of all claims data. An example is given below.

### 3.2. Credibility

It is well known that credibility estimators perform poorly in the presence of large claims (see Bühlmann et al. (1982)).

As mentioned in section 2.2 the credibility estimator is equal to the posterior mean in the case of likelihoods belonging to the simple exponential family with natural conjugate priors. Since this models are not outlier prone the above statement is hardly surprising. This situation has led to different ad hoc adjustments of credibility estimators. Gisler (1980) combines credibility procedures with data trimming. It is interesting to note that in the case of a Laplace likelihood and a normal prior, the posterior mean also leads to data trimming. Künsch (1992) advocates the replacement of the empirical mean in the credibility formula by a robust estimator of the unknown mean. Gisler and Reinhard (1993) propose a specific procedure for the data structure presented in section 3.1. Below we shall apply this procedure to a practical example.

## 4. EXAMPLE

The following claims data stem from the Swiss Association of Property Insurers. (For confidentiality reasons claims intensities have all been multiplied by the same undisclosed factor.)

Risk Category Category	Year	Sum Insured in CHF 1'000.-	Claims Intensity in ‰
1. Stone Industry	1	8'952'537	1.170
	2	9'408'941	0.923
	3	9'116'202	0.790

<b>Risk Category</b> <b>Category</b>	<b>Year</b>	<b>Sum Insured</b> <b>in CHF 1'000.-</b>	<b>Claims Intensity</b> <b>in ‰</b>
	4	9'233'632	0.494
	5	9'341'821	1.405
2. Metal Industry	1	54'637'719	1.299
	2	56'197'669	0.592
	3	56'014'549	0.640
	4	54'660'986	2.863
	5	57'393'239	0.446
3. Wood Industry	1	6'039'217	2.844
	2	6'217'858	2.337
	3	5'770'074	2.907
	4	4'961'525	2.396
	5	5'209'193	0.972
4. Paper Industry	1	15'031'003	1.468
	2	15'862'988	1.570
	3	16'637'453	0.322
	4	16'474'230	0.556
	5	15'962'600	6.329
5. Textile Industry	1	7'690'266	0.464
	2	7'817'476	1.601
	3	8'489'434	2.175
	4	8'298'066	0.802
	5	7'810'418	0.181
6. Food Industry	1	13'518'262	1.122
	2	14'101'545	0.985
	3	13'027'446	0.763
	4	12'654'978	0.395
	5	12'395'113	0.564
7. Chemical Industry	1	18'033'514	0.801
	2	19'599'797	1.702
	3	23'505'751	0.174
	4	16'665'459	0.250
	5	11'548'235	0.308
8. Energy	1	21'969'611	0.466
	2	23'257'289	0.413
	3	21'524'998	0.369
	4	21'390'824	0.194
	5	23'346'584	0.251
9. Shops and Hotels	1	44'119'033	0.544
	2	45'321'074	0.411
	3	43'405'903	0.583
	4	43'309'859	0.790
	5	41'759'826	0.601
<b>Total</b>	1	189'991'162	0.990
	2	197'784'637	0.857
	3	197'491'810	0.663
	4	187'649'559	1.260
	5	184'767'029	1.016

The data consist of the claims intensities (claim amount divided by sum insured) from nine different categories of risks recorded during a period of five consecutive years. The claim amounts pertain to fire claims and exclude claims from elemental perils. This latter category must be analyzed separately since it introduces a positive correlation between categories of risks in any given year. The claims intensities are of particular interest since they define the rate to be applied to the sum insured in order to cover the expected loss costs (i.e. the pure risk rates). The data provide a good example of the typical problems with which the practitioner is confronted when developing a tariff. Most of the risk categories are too small to enable a pricing based on the experience of the individual category only. Even the largest risk category, the metal industry, represents only approximately one quarter of the total sum insured of all risk categories. It will therefore also benefit from an appropriate pooling of the experience of all risk categories. Credibility theory seems to be the solution to this problem since it combines the experience of the individual risk category and of the collective. For many risk categories, the individual experience varies strongly from year to year as extraordinary losses occur or fall out of the statistics of the last five years. This leads to undesirable, strong fluctuations in pure risk premiums, a problem which is well known to practitioners. A robust version of credibility seems to be the solution to this problem.

We analyze the data with a standard credibility procedure and use the robust credibility procedure proposed by Gisler and Reinhard (1993). We then perform a fully fledged Bayesian analysis, based on the model of section 3.1. Finally we compare the performance of the different methods in this particular case. In order to analyze the data, we introduce a similar notation to the one used in section 3.1.

- $x_{ij}$  denotes the claims intensity from risk category number  $i$  in year  $j$  ( $i = 1, \dots, 9, j = 1, \dots, 5$ ).  $x_{ij}$  is expressed in ‰ of the sum insured.  
 $v_{ij}$  denotes the corresponding sum insured or volume measure attached to  $x_{ij}$ .  $v_{ij}$  is expressed in billions CHF.

Let  $n = 5$  and  $N = 9$  denote the number of years and risk categories respectively. The individual premium of risk category number  $i$  is

$$\bar{x}_i = \sum_{j=1}^n \frac{v_{ij}}{v_i} x_{ij} \quad \text{with} \quad v_i = \sum_{j=1}^n v_{ij}.$$

According to Bühlmann and Straub (1970), the credibility premium of risk category number  $i$  is

$$\hat{\mu}_i = \mu + \alpha_i (\bar{x}_i - \mu)$$

where  $\mu$  is the portfolio mean and  $\alpha_i$  is the credibility factor pertaining to risk number  $i$

$$\alpha_i = \frac{v_i \cdot b}{v_i \cdot b + w}.$$

$b$  and  $w$  are the between risks and within risk variance components respectively.

Bühlmann and Straub also propose ways to estimate  $\mu, b$  and  $w$  from the sample

$$\begin{aligned}\hat{\mu} &= \sum_{i=1}^N \frac{\alpha_i}{\alpha} \bar{x}_i, \quad \text{with } \alpha = \sum_{i=1}^N \alpha_i \\ \hat{w} &= \frac{1}{N} \frac{1}{n-1} \sum_{i=1}^N \sum_{j=1}^n v_{ij} (x_{ij} - \bar{x}_i)^2 \\ \hat{b} &= c^{-1} \left\{ \sum_{i=1}^N \frac{v_i}{v} (\bar{x}_i - \bar{x})^2 - (N-1) \frac{\hat{w}}{v} \right\}\end{aligned}$$

where

$$\begin{aligned}v &= \sum_{i=1}^N v_i, \\ \bar{x} &= v^{-1} \sum_{i=1}^N v_i \bar{x}_i, \\ c &= \sum_{i=1}^N \frac{v_i}{v} \left( 1 - \frac{v_i}{v} \right)\end{aligned}$$

The estimator of the portfolio mean and of the two variance components are

$$\begin{aligned}\hat{\mu} &= 0.981 \cdot 10^{-3} \\ \hat{w} &= 19.162 \cdot 10^{-6} \\ \hat{b} &= 0.108 \cdot 10^{-6}\end{aligned}$$

Note that  $\hat{\mu}$  is different from the average premium

$$\bar{x} = 0.953 \cdot 10^{-3}$$

The individual premiums  $\bar{x}_i$  and the credibility premiums  $\hat{\mu}_i$  are tabulated below.

Gisler and Reinhard (1993) propose a robust version of credibility analysis. They assume that the individual mean  $\mu(\theta_i) = E[x_{ij} | \theta_i]$  is the sum of an ordinary part and of an excess-part

$$\mu(\theta_i) = \mu_0(\theta_i) + \mu_{xs}$$

whereby the excess part is independent of the individual risk category. The ordinary part  $\mu_0(\theta_i)$  is estimated based on robust statistics  $t_i$ , and  $\mu_{xs}$  is estimated based on the observed  $xs$  claims.

As robust estimator, Gisler and Reinhard propose a so called  $M$ -estimator  $t_i$ , which is obtained as the solution of the following implicit equation

$$t_i = \sum_{j=1}^n \frac{v_{ij}}{v_i} \min(x_{ij}, c_{ij} t_i)$$

$$\text{where } c_{ij} = 1 + \left( \frac{\bar{v}}{v_{ij}} \right)^{\frac{1}{2}}$$

$$\bar{v} = \frac{1}{n \cdot N} \sum_{i=1}^N \sum_{j=1}^n v_{ij}$$

$t_i$  has the particularity that the influence of any observation is never larger than that of a zero observation, i.e. 'large' observations are truncated with the truncation point depending on the individual risk category. The robust credibility estimator proposed by Gisler and Reinhard is of the following form

$$\hat{\mu}_i = \hat{\mu}_t + \alpha_i (t_i - \hat{\mu}_t) + \hat{\mu}_{xs}$$

where

$$\hat{\mu}_{xs} = \frac{1}{v} \sum_{i=1}^N \sum_{j=1}^n v_{ij} \cdot x s_{ij}$$

$$x s_{ij} = x_{ij} - t_{ij}$$

$$t_{ij} = \min(x_{ij}, c_{ij} \cdot t_i)$$

$$\alpha_i = \frac{v_{i \cdot} \cdot b_t}{v_{i \cdot} \cdot b_t + w_t}$$

$$\hat{\mu}_t = \sum_{i=1}^N \frac{\alpha_i}{\alpha} t_i, \quad \alpha = \sum_{i=1}^N \alpha_i$$

$b_t$  and  $w_t$  are obtained from a modified version of the estimators proposed by Bühlmann Straub (see Gisler and Reinhard, 1993).

Using the above formulae we have obtained

$$\underline{t} = (0.956, 0.871, 2.320, 1.349, 1.063, 0.776, 0.532, 0.339, 0.584) \cdot 10^{-3}$$

and it is seen that with the exception of risk category 2, 4 and 7 we have  $t_i = \bar{x}_i$ . The contribution from excess claims is

$$\hat{\mu}_{xs} = 0.152 \cdot 10^{-3}$$

The variance components pertaining to ordinary claims are

$$\begin{aligned}w_t &= 10.885 \cdot 10^{-6} \\ b_t &= 0.061 \cdot 10^{-6}\end{aligned}$$

The portfolio mean pertaining the ordinary claims is

$$\hat{\mu}_t = 0.836 \cdot 10^{-3}$$

The robust credibility premiums are tabulated below.

It is interesting to note that the ratio of within to between variance is approximately the same in the Bühlmann Straub model

$$\frac{w}{b} = \frac{19.2}{0.108} = 177.8$$

and in the Gisler Reinhard Model

$$\frac{w_t}{b_t} = \frac{10.9}{0.061} = 178.7.$$

This means that the credibility factors of a given risk category are approximately the same in both models.

We now turn to the fully fledged Bayesian model. We assume that for  $i = 1, \dots, N$  and  $j = 1, \dots, n$

$$\tilde{x}_{ij} | \mu_i, \tau_{ij}$$

are independent and gamma distributed with mean  $\mu_i$  and variance  $\frac{\tau_{ij}}{\nu_{ij}}$ .

The corresponding density is found in section 3.1. As in section 3.1 we assume that  $\mu_i (i = 1, \dots, N)$  and  $\tau_{ij} (i = 1, \dots, N, j = 1, \dots, n)$  are all independent and gamma distributed with

$$\begin{aligned}E(\tilde{\mu}_i) &= \mu, & Var(\tilde{\mu}_i) &= b, \\ E(\tilde{\tau}_{ij}) &= w, & Var(\tilde{\tau}_{ij}) &= t,\end{aligned}$$

As in the credibility model we estimate  $m, b$  and  $w$  based on the sample. We give  $t$  a 'very large' value to ensure that  $\bar{\tau}_{ij}$  gets largely determined by the data.

The estimators proposed by Bühlmann Straub are unbiased and we have used the same values  $\hat{\mu}$ ,  $\hat{w}$  and  $\hat{b}$  as in the standard credibility model above. We have set  $t = 10'000$ .

For each  $i$  ( $i = 1, \dots, N$ ), the least square estimator of  $\mu_i$  is

$$E(\tilde{\mu}_i | D) = \frac{\int \mu_i \left[ \prod_{i=1}^n f(x_{ij} | \mu_i, \tau_{ij}) f(\tau_{ij}) \right] f(\mu_i) d\mu_i d\tau_{i_1}, \dots, d\tau_{i_n}}{f(x_{i1}, x_{i2}, \dots, x_{in})}$$

In this case  $n = 5$  and the above integration is six dimensional. A six dimensional numerical integration is extremely cumbersome and we have derived the above quantities through Markov Chain Monte Carlo methods. For details on the method used see Appendix 3. The posterior means are tabulated below.

The results of the analysis are found in the following table, which displays the pure risk premiums of the different risk categories computed according to the methods described above

	Risk Category								
	1	2	3	4	5	6	7	8	9
Portfolio Mean [‰]	0.981	0.981	0.981	0.981	0.981	0.981	0.981	0.981	0.981
Individual Premium [‰]	0.956	1.155	2.320	2.032	1.063	0.776	0.667	0.339	0.584
Credibility Premium [‰]	0.976	1.088	1.165	1.308	0.996	0.925	0.876	0.733	0.762
Robust Credibility Premium [‰]	1.013	1.010	1.190	1.147	1.030	0.973	0.888	0.798	0.850
Posterior Mean [‰]	1.03	0.80	1.61	1.30	1.13	0.91	0.77	0.63	0.69

Looking at the results of the analysis leads to the following remarks

#### **Risk Category 1, Stone Industry**

Credibility premium (0.98), robust credibility premium (1.01), posterior mean (1.03) individual experience (0.96) and portfolio mean (0.98) are all close.

#### **Risk Category 2, Metal Industry**

Credibility premium (1.09) is higher than robust credibility premium (1.01) and much higher than posterior mean (0.80). This is due to the treatment of the one outlier in the claims experience (2.863). It is fully taken into account by the credibility premium, truncated by the robust credibility premium and nearly discarded by the posterior mean.

#### **Risk Category 3, Wood Industry**

Credibility premium (1.17) and robust credibility premium (1.19) are both much lower than posterior mean (1.61) which gives more weight to the individual experience of the risk (individual premium 2.32).

#### **Risk Category 4, Paper Industry**

The credibility premium (1.31) is considerably higher than the robust credibility premium (1.15). This is due to the fact that the exceptionally large claims intensity of the fifth year is truncated by the robust credibility premium. The posterior mean (1.30) is approximately equal to the credibility premium. This is so

because the posterior mean weighs down both the exceptionally large claims intensity of the fifth year, and the exceptionally small claims intensity of the third year

### **Risk Category 5, Textile Industry**

The credibility premium (1.00) and the robust credibility premium (1.03) are approximately the same. The posterior mean (1.13) is higher due to the fact that this estimator weighs down the exceptionally low claims experience of the fifth year.

### **Risk Category 6, Food Industry**

All three estimators produce similar results.

### **Risk Category 7, Chemical Industry**

The two credibility estimators produce results which are practically identical (0.88 and 0.89). The posterior mean (0.77) is much closer to the individual premium (0.67). The posterior mean gives more weight to the individual experience of the risk.

### **Risk Category 8, Energy**

Same remark as for risk category 7. In addition one notices the higher price charged by the robust credibility estimator (0.80), compared to the credibility estimator (0.73) which is due to the surcharge for  $x_s$  risks which is spread across all risk categories.

### **Risk Category 9, Shops and Hotels**

Same remark as for risk category 8.

The full Bayesian analysis gives more weight to the experience of individual risk categories while at the same time weighing down both exceptionally large and exceptionally small claims intensities. The posterior mean yields pure risk premiums which vary more widely than the risk premiums produced by any of the two credibility methods. In practical applications price ranges are more in line with credibility results than with the results of the fully fledged Bayesian analysis, i.e. they are more narrow. Practitioners however are aware of the fact that good risk categories are more profitable than average risk categories — in spite of the price differentials encountered in the market —. They endeavour to write many of those risks. Practitioners also know that bad risk categories are less profitable than average risk categories — in spite of market price differentials. They try to avoid those risks. Practitioners know that price ranges encountered in the market are too narrow and take this into account through qualitative underwriting considerations (write few high hazard risks and many low hazard risks). A more rational pricing would make this unnecessary.

The posterior mean has the advantage to weigh down both unusually low claims intensities and unusually high claims intensities.

The robust credibility estimator only takes corrective action for unusually high claims intensities.



The posterior mean presented here has the particularity to discard very large outlying observations. This is a drawback in an insurance context, since after a certain threshold, the higher the unusual claims intensity, the lower the pure risk premium. This drawback can probably be corrected by an appropriate choice of the prior distribution of  $\bar{\tau}$ , the unknown variance of the individual claims intensity, as is shown in the case of symmetric distributions.

Both, in the case of credibility and robust credibility, we have a mathematical allocation of the total claims amount, in the sense that

$$\sum_{j=1}^N \hat{\mu}_i \frac{v_{i.}}{v} = \bar{x}.$$

In the case of the robust Bayes method presented here this is only true in the expectation, i.e.

$$E(E(\tilde{\mu}_i | D)) = m.$$

However it is not felt that this is a major drawback, since the claims amount of the past is the realization of a random variable and is generally not equal the expected claims amount of the future, which is the quantity of interest.

All three estimators, credibility, robust credibility and posterior mean require a fair amount of computational work since the prior mean and the two variance components must be estimated from the data, which is a cumbersome exercise. In the case of the full Bayesian analysis, one has to also perform a lengthy Monte Carlo simulation to obtain the posterior mean of each risk category

## REFERENCES

- AITCHISON, J. and DUNSMORE I.R. (1975) *Statistical Prediction Analysis*, Cambridge University Press.
- BERGER, J. (1993) An Overview of Robust Bayesian Analysis. Technical Report, Department of Statistics, Purdue University.
- BÜHLMANN, H. and STRAUB E. (1970) Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **70**(1), 1970.
- BÜHLMANN, H., GISLER, A., JEWELL, W.S. (1982), Excess Claims and Data Trimming in the Context of Credibility Rating Procedures, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **1**, 1982.
- GILKS, W.R., RICHARDSON, S. and SPIEGELHALTER, D.J. (1996) *Markov Chain Monte Carlo in Practice*, Chapman and Hall, 1996.
- GISLER, A. (1980) Optimum Trimming of Data in the Credibility Model, *Mitteilungen der Schweiz. Vereinigung der Versicherungsmathematiker*, **3**, 1980.
- GISLER, A. and REINHARD, P. (1993) Robust Credibility, *ASTIN Bulletin*, **23**(1).
- JEWELL, W.S. (1974) Credible Means are Exact Bayesian for Exponential Families, *ASTIN Bulletin*, Vol. **VIII**.
- KÜNSCH, H.R. (1992) Robust Methods for Credibility, *ASTIN Bulletin*, **22**(1).
- O'HAGAN, A. (1994) *Kendalls Advanced Theory of Statistics*, Vol. 2B, Bayesian Inference.
- SCHNIEPER, R. (1993) Praktische Erfahrungen mit Grossschadenverteilungen, *Mitteilungen der Schweiz. Vereinigung der Versicherungsmathematiker*, Heft 2, 1993.

## APPENDIX 1

**Theorem****Assumptions**

- (i) Given  $\theta$ , the claims are independent and Laplace distributed with mean  $\theta$  and precision  $q$

$$f(x|\theta) = \sqrt{\frac{q}{2}} e^{-\sqrt{2q}|x-\theta|}$$

- (ii) The unknown mean  $\tilde{\theta}$  is normally distributed with mean  $m$  and precision  $p$

$$f(\theta) = \sqrt{\frac{p}{2\pi}} e^{-\frac{p}{2}(\theta-m)^2}$$

The following statements hold true:

- (i)  $\lim_{x \rightarrow \pm\infty} E[\tilde{\theta}|x] = E[\tilde{\theta}] \pm \frac{\sqrt{2q}}{p}$
- (ii)  $E[\tilde{\theta}|x]$  is an increasing function of  $x$ .

**Proof**

$$E[m(\tilde{\theta})|x] = \frac{\int_{-\infty}^{\infty} \theta e^{-\sqrt{2q}|x-\theta| - \frac{p}{2}(\theta-m)^2} d\theta}{\int_{-\infty}^{\infty} e^{-\sqrt{2q}|x-\theta| - \frac{p}{2}(\theta-m)^2} d\theta}$$

without loss of generality we set  $m=0$  which amounts to a choice of the origin and obtain after some straightforward simplifications

$$E[m(\tilde{\theta})|x] = \frac{\int_{-\infty}^x \theta e^{+\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta + e^{+2\sqrt{2q}x} \int_x^{\infty} \theta e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta}{\int_{-\infty}^x e^{+\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta + e^{+2\sqrt{2q}x} \int_x^{\infty} e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta}$$

on the other hand

$$e^{2\sqrt{2q}x} \int_x^{\infty} \theta e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta \leq e^{2\sqrt{2q}x} \int_x^{\infty} \theta e^{-\frac{p}{2}\theta^2} d\theta = e^{2\sqrt{2q}x} \frac{1}{p} e^{-\frac{p}{2}x^2}$$

hence the above term goes to 0 for  $x \rightarrow \infty$  and

$$e^{2\sqrt{2q}x} \int_x^\infty e^{-\sqrt{2q}\theta} e^{-\frac{p}{2}\theta^2} d\theta \leq e^{2\sqrt{2q}x} \int_x^\infty \theta e^{-\frac{p}{2}\theta^2} d\theta$$

for  $x \geq 1$ , and it follows that this term also goes to 0 for  $x \rightarrow \infty$ . Hence

$$\begin{aligned} \lim_{x \rightarrow \infty} E[m(\tilde{\theta})|x] &= \lim_{x \rightarrow \infty} \frac{\int_x^\infty \theta e^{\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta}{\int_{-\infty}^x e^{\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta} \\ &= \frac{\int_{-\infty}^\infty \theta e^{-\frac{p}{2}\left(\theta - \frac{\sqrt{2q}}{p}\right)^2} d\theta}{\int_{-\infty}^\infty e^{-\frac{p}{2}\left(\theta - \frac{\sqrt{2q}}{p}\right)^2} d\theta} \\ &= \frac{\sqrt{2q}}{p} \end{aligned}$$

And a similar exercise for  $x \rightarrow -\infty$  proves statement (i) of the theorem.

We now turn to the proof of statement (ii).

$$\begin{aligned} E[m(\tilde{\theta})|x] &= \frac{\int_{-\infty}^x \theta e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta + \int_x^\infty \theta e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta}{\int_{-\infty}^x e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta + \int_x^\infty e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta} \\ \frac{d}{dx} E[m(\tilde{\theta})|x] &= \frac{u'v - uv'}{v^2} \end{aligned}$$

with

$$\begin{aligned} u' &= xe^{-\frac{p}{2}x^2} - \sqrt{2q} \int_{-\infty}^x \theta e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta \\ &\quad - xe^{-\frac{p}{2}x^2} + \sqrt{2q} \int_{-\infty}^x \theta e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \\ &= -\sqrt{2q} \int_{-\infty}^x \theta e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta + \sqrt{2q} \int_x^\infty \theta e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \end{aligned}$$

$$\begin{aligned}
&= \sqrt{2q}(-A_1 + A_2) \\
v &= \int_{-\infty}^x e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta + \int_x^{\infty} e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \\
v &= B_1 + B_2 \\
u &= A_1 + A_2 \\
v' &= e^{-\frac{p}{2}x^2} - \sqrt{2q} \int_{-\infty}^x e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta \\
&\quad - e^{-\frac{p}{2}x^2} + \sqrt{2q} \int_x^{\infty} e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \\
&= \sqrt{2q}(-B_1 + B_2)
\end{aligned}$$

Hence

$$\begin{aligned}
u'v - uv' &= \sqrt{2q}((-A_1 + A_2)(B_1 + B_2) - (A_1 + A_2)(-B_1 + B_2)) \\
u'v - uv' &= 2\sqrt{2q}(-A_1B_2 + A_2B_1)
\end{aligned}$$

$$\begin{aligned}
A_2B_1 - A_1B_2 &= \int_x^{\infty} \theta e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \int_{-\infty}^x e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta \\
&\quad - \int_{-\infty}^x \theta e^{-\sqrt{2q}(x-\theta) - \frac{p}{2}\theta^2} d\theta \int_x^{\infty} e^{-\sqrt{2q}(\theta-x) - \frac{p}{2}\theta^2} d\theta \\
&= \int_x^{\infty} \theta e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta \int_{-\infty}^x e^{+\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta \\
&\quad - \int_{-\infty}^x \theta e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta \int_x^{\infty} e^{-\sqrt{2q}\theta - \frac{p}{2}\theta^2} d\theta \\
&= e^{\frac{q}{p}} \int_x^{\infty} \theta e^{-\frac{p}{2}\left(\theta + \frac{\sqrt{2q}}{p}\right)^2} d\theta \int_{-\infty}^x e^{-\frac{p}{2}\left(\theta - \frac{\sqrt{2q}}{p}\right)^2} d\theta \\
&\quad - e^{\frac{q}{p}} \int_{-\infty}^x \theta e^{-\frac{p}{2}\left(\theta - \frac{\sqrt{2q}}{p}\right)^2} d\theta \int_x^{\infty} e^{-\frac{p}{2}\left(\theta + \frac{\sqrt{2q}}{p}\right)^2} d\theta
\end{aligned}$$

We introduce the following random variable

$$\begin{aligned}
X_- &\sim N\left(\mu = -\frac{\sqrt{2q}}{p}, \sigma^2 = p^{-1}\right) \\
X_+ &\sim N\left(\mu = \frac{\sqrt{2q}}{p}, \sigma^2 = p^{-1}\right)
\end{aligned}$$

We obtain

$$\begin{aligned}
 A_2 B_1 - A_1 B_2 &= e^{\frac{q}{p}} (E[X_-; X_- > x] \cdot \Pr[X_+ < x] \\
 &\quad - E[X_+; X_+ < x] \cdot \Pr[X_- > x]) \\
 &= e^{\frac{q}{p}} \Pr[X_+ < x] \cdot \Pr[X_- > x] \\
 &\quad \cdot (E[X_- | X_- > x] - E[X_+ | X_+ < x]) \\
 &> 0
 \end{aligned}$$

And it is seen that  $\frac{d}{dx} E[m(\tilde{\theta}) | x] > 0$

which proves statement (ii) of the theorem.

## APPENDIX 2

As in section 3.1, we assume that the claims data is distributed according to the following density

$$f(x|\mu, \tau) = \frac{\left(\frac{\mu}{\tau}\right)^{\frac{\mu^2}{\tau}}}{\Gamma\left(\frac{\mu^2}{\tau}\right)} = e^{-\frac{\mu}{\tau}x} \cdot x^{\frac{\mu^2}{\tau}-1}$$

i.e. it is conditionally gamma distributed given the mean  $\mu$  and the variance  $\tau$ . We assume that the unknown prior mean  $\mu$  and prior variance  $\tau$  are also gamma distributed.

$$\begin{aligned} f(\mu) &= k \cdot \mu^{\frac{m^2}{b}-1} \cdot e^{-\frac{m}{b}\mu} \\ f(\tau) &= k \cdot \tau^{\frac{w^2}{t}-1} \cdot e^{-\frac{w}{t}\tau} \end{aligned}$$

We make the following assumptions about the hyperparameters

$$\begin{aligned} E(E(\tilde{x}|\mu, \tau)) &= E(\tilde{\mu}) = m = 1 \\ E(Var(\tilde{x}|\mu, \tau)) &= E(\tilde{\tau}) = w = 2 \\ Var(E(\tilde{x}|\mu, \tau)) &= Var(\tilde{\mu}) = b = 2 \end{aligned}$$

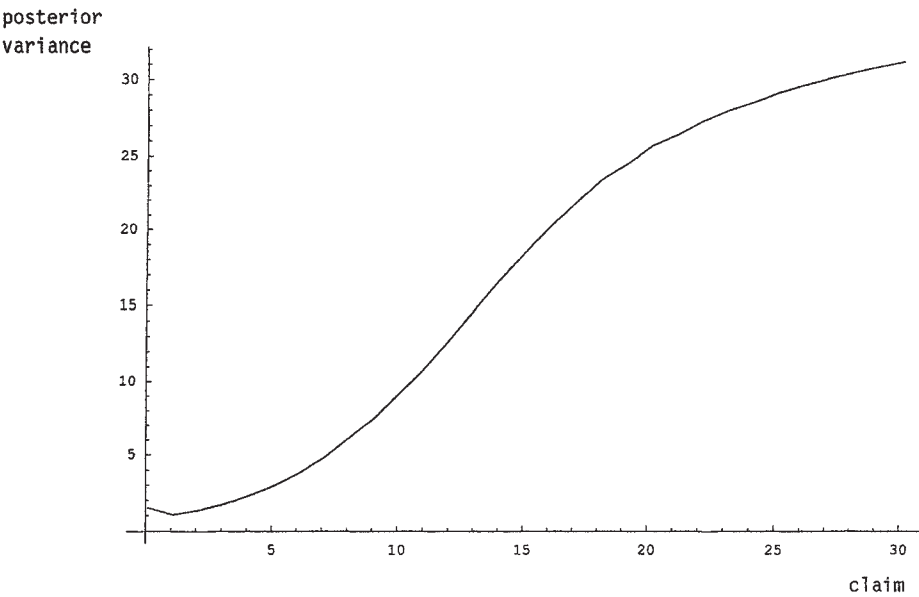
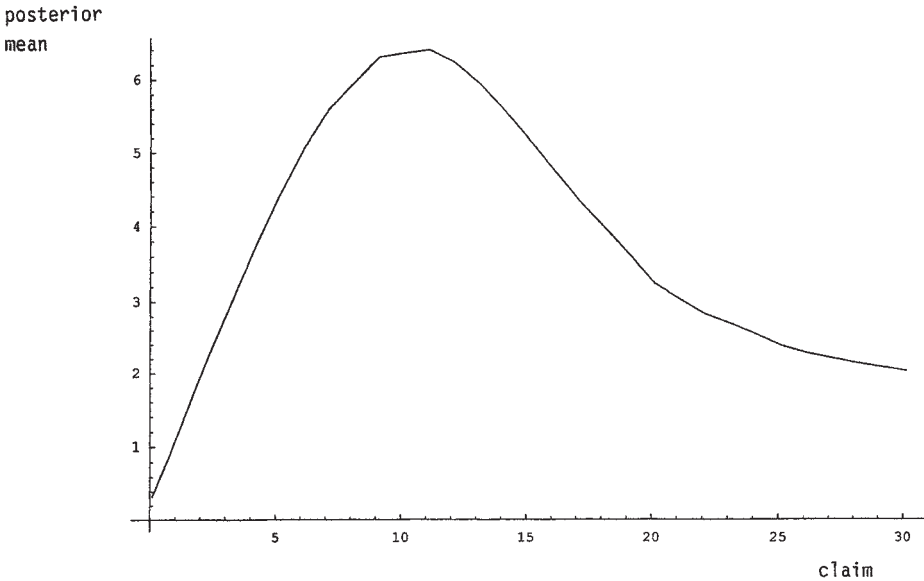
and finally

$$Var(\tilde{\tau}) = t = 100$$

which means that  $\tau$  is essentially determined by the data.

Based on the above assumptions we have derived the posterior mean  $E(\tilde{\mu}|x)$  and the posterior variance  $E(\tilde{\tau}|x)$  as a function of the claims data  $x$ . The derivation has been performed through numerical integration. The two functions are shown below.

As is seen from the graphs, the posterior mean peaks at approximately  $x = 10$  and decreases sharply thereafter. Whilst the data is not totally discarded, it is given a very low weight as it increases beyond say  $x = 20$ . This is confirmed by the analysis of the graph of the posterior variance, which increases dramatically in the range between approximately  $x = 10$  and  $x = 20$ . The posterior mean is thus robust, since it dramatically weighs down ‘large’ outlying observations.



## APPENDIX 3

**Markov Chain Monte Carlo**

We consider risk category number  $i$ . We are interested in the posterior mean of the unknown parameters

$$\mu_i, \tau_{i1}, \tau_{i2}, \dots, \tau_{i5}.$$

Our primary focus is on the posterior mean of  $\mu_i$  i.e.  $E(\mu_i | D)$ .  $\tau_{i1}, \tau_{i2}, \dots, \tau_{i5}$  are ancillary statistics. They are only interesting in so far as they give information about the influence of  $x_{i1}, x_{i2}, \dots, x_{i5}$  on  $E(\mu_i | D)$ .

According to Bayes' Theorem, and to the assumptions in section 4, the posterior density of the parameters

$$f(\mu_i, \tau_{i1}, \tau_{i2}, \dots, \tau_{i5} | D)$$

is proportional to

$$\left[ \prod_{j=1}^n (f(x_{ij} | \mu_i, \tau_{ij}) \cdot f(\tau_{ij})) \right] \cdot f(\mu_i)$$

Thereby  $D = (x_{11}, x_{12}, \dots, x_{Nn})$  denotes the data. It is seen, that the computation of the posterior mean

$$E(\mu_i | D)$$

involves a six-dimensional numerical integration.

Another possibility to evaluate  $E(\mu_i | D)$  consists in generating a large sample  $(\mu_i^1, \mu_i^2, \dots, \mu_i^s)$  from  $f(\mu_i, \tau_{i2}, \dots, \tau_{i5} | D)$  and in approximating

$$E(\mu_i | D) \simeq \frac{1}{s} \sum_{t=1}^s \mu_i^t$$

where  $t$  is the index numbering individual sample elements, and  $s$  is the sample size which depends on the desired precision of the estimator.

A possible way of generating such a sample is through a Markov Chain having  $f(\mu_i, \tau_{i1}, \tau_{i2}, \dots, \tau_{i5} | D)$  as its stationary distribution. The Metropolis-Hastings algorithm (see W.R. Gilks et al., 1996) provides a way to construct such a sample. At each time  $t$ , the next state is chosen by first sampling a candidate point

$$Y = (\mu_i^{t+1}, \tau_{i1}^{t+1}, \dots, \tau_{i5}^{t+1})$$

from a proposal distribution

$$p(Y | \mu_i^t, \tau_{i1}^t, \dots, \tau_{i5}^t)$$



which may depend on the current point

$$X_t = (\mu_t^t, \tau_{i1}^t, \dots, \tau_{i5}^t).$$

The candidate point is then accepted with probability

$$\alpha(X_t, Y) = \min\left(1, \frac{f(Y|D) \cdot p(X_t|Y)}{f(X_t|D) \cdot p(Y|X_t)}\right)$$

If the candidate is accepted, the next state becomes  $X_{t+1} = Y$ . If the candidate is rejected, the chain does not move, i.e.  $X_{t+1} = X_t$ . Remarkably, under very general conditions, the stationary distribution of the chain is  $f(\cdot|D)$ . For a proof of the statement and a discussion on the speed of convergence of the distribution of  $X_t$  to the stationary distribution see Gilks et al., 1996.

As a proposal distribution we have chosen the product of the following marginal distributions

$$\begin{aligned} p(\mu_i, \tau_{i1}, \tau_{i2}, \tau_{i3}, \tau_{i4}, \tau_{i5}) = & f(\mu_i, \tau_0, \dots, \tau_0 | D). \\ & f(\mu_0, \tau_{i1}, \tau_0, \dots, \tau_0 | D). \\ & f(\mu_0, \tau_0, \tau_{i2}, \tau_0, \dots, \tau_0 | D). \\ & \dots\dots\dots \\ & f(\mu_0, \tau_0, \tau_0, \dots, \tau_0, \tau_{i5} | D). \end{aligned}$$

where  $\mu_0 = 0.981$  and  $\tau_0 = 19.162$  are the a priori mean and the within variance estimated from the data. Note that in our case  $p(Y|X_t) = p(Y)$  does not depend on the present state. As a starting point for the Markov Chain, we have chosen the a priori mean and the within variance component estimated from the data  $\mu_0 = 0.981$ ,  $\tau_0 = 19.162$ .

For a proof of the convergence of the distribution of  $X_t$  to the stationary distribution see Wilks et al., 1996. The convergence of

$$\frac{1}{s} \sum_{t=1}^s \mu_i^t$$

to  $E(\mu_i|D)$  is ensured by the ergodic theorem.

The speed of convergence has been checked empirically for a simplified model with only two observations per risk category. In such a case there are only three parameters to estimate  $(\mu_i, \tau_{i1}, \tau_{i2})$  which leads to a three dimensional numerical integration. This operation can be performed with a reasonable amount of computational time. It was found that the precision of the Monte Carlo estimator of the posterior mean was acceptable for  $s = 10'000$  and good for  $s = 50'000$ . The precision of the estimator of the within variances was acceptable for  $s = 100'000$ . Thereby we have used the following approximation formulae

$$\hat{\tau}_{ij} = E(\tau_{ij}|D) = \frac{1}{s} \sum_{t=1}^s \tau_{ij}^t \quad j = 1, \dots, 2.$$

As a consequence two runs of the Monte Carlo estimator with  $N = 50'000$  were performed for each risk category. The differences in estimates between the two runs give an idea of the precision of the estimators. As final estimates we have taken the arithmetic mean of the results of the two runs. As an example we give the result of the two runs performed for risk category number 1

$$E(\mu_i|D) \quad \frac{\text{run 1}}{1.029} \quad \frac{\text{run 2}}{1.024} \quad \frac{\text{average}}{1.03}$$

where the result is only recorded up to the first two decimals because of the limited precision of the estimator.

$j:$	1	2	3	4	5
$x_{1j}$	1.170	0.923	0.790	0.494	1.405
$\hat{\tau}_{1j}^1$	5.3	2.9	7.5	18.7	9.8
$\hat{\tau}_{1j}^2$	5.8	2.7	7.9	18.4	9.5
$\hat{\tau}_{1j}$	6	3	8	19	10

where  $\hat{\tau}_{1j}^1$ ,  $\hat{\tau}_{1j}^2$  and  $\hat{\tau}_{1j}$  denote the result of the first and second run and the average of the two runs respectively. We do not record the first decimal of the average because of the limited precision of the estimator.

The following table provides an overview over the results for all risk categories.

$i$	$E(\mu_i D)$	$\hat{\tau}_{i1}$	$\hat{\tau}_{i2}$	$\hat{\tau}_{i3}$	$\hat{\tau}_{i4}$	$\hat{\tau}_{i5}$
1	1.03	6	3	8	19	10
2	0.80	21	20	18	112	49
3	1.61	17	17	24	13	5
4	1.30	11	12	57	38	115
5	1.13	23	12	19	11	40
6	0.91	6	1	8	21	14
7	0.77	5	31	58	30	13
8	0.63	14	6	20	36	29
9	0.69	14	40	10	19	13

RENÉ SCHNIEPER  
Converium  
General Guisan Quai 26  
CH-8022 Zurich, Switzerland  
E-mail: [rene.schnieper@converium.com](mailto:rene.schnieper@converium.com)